

Texture Classification Using Wavelet packet and Fourier Transforms

Xiaou Tang and W. Kenneth Stewart

Deep Submergence Laboratory

Department of Applied Ocean Physics and Engineering

Woods Hole Oceanographic Institution

Woods Hole, MA 02543

Department of Ocean Engineering

Massachusetts Institute of Technology

Cambridge, MA 02139

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

Abstract —A new texture classification algorithm using wavelet packet transform is proposed. It uses principal component analysis technique and statistical distance measurement to combine and select frequency channel features to give improved classification performance. Comparison is also made between wavelet packet transform features and Fourier transform features on a set of eight optical texture images with several level of white noise added. Both algorithms are successfully applied to the classification of under-ice sidescan sonar images.

I. INTRODUCTION

Texture analysis has found many important applications in such areas as medical imaging, computer vision, and remote sensing. Many successful algorithms have been proposed over the last few decades. Recently, multichannel analysis methods, including texture energy measurement [1], the eigenfilter method [2], and Gabor filter method [3], have been repetitively proved to perform better than other techniques. The newly developed wavelet analysis technique [4] [5] provides yet another useful framework for multiscale image processing. The texture research community is currently devoting considerable effort to wavelet applications in texture analysis. Henke-Reed and Cheng [6] applied wavelet transforms to texture images, using the energy ratios between frequency channels as features. Chang and Kuo [7] proposed a tree-structured wavelet transform algorithm for texture classification, which is similar to the wavelet packet best bases selection algorithm of Coifman [5]. Laine and Fan [9] used the wavelet

packet transform energy measurements directly as texture features in their texture classification approach.

These researchers have demonstrated that the wavelet transform is a valuable tool for texture analysis. However, a common problem with these approaches is that they are all direct applications of existing wavelet processing algorithms, which are ideal for signal representation but not necessarily the best for signal discrimination. To fully utilize the power of a wavelet packet transform, new techniques tailored for extracting features of greater discrimination ability must be developed. In this paper we propose the use of principal component analysis technique and statistical distance measurement to combine and select frequency-channel features that give improved classification performance.

We also compare this new approach with the Fourier transform texture classification method which we proposed in [9]. Just as the ideal tool for nonstationary signal analysis is a wavelet transform, the ideal tool for stationary signal is a Fourier transform. Since texture signals are mostly stationary, we should expect the Fourier transform to generate better results.

The new algorithms are tested on two data sets. The first includes eight types of natural optical images obtained from the MIT Media Lab Vistex texture data base. We hope to get more conclusive result from this larger classes of data. For our real world application, we apply the algorithms on a second set of sidescan sonar images from a sonar survey of an Arctic under-ice canopy [10].

This work is supported by the Office of Naval Research

We first give a brief review of the wavelet transform and wavelet packet transform in section II.

19970430 019

DTIC QUALITY INSPECTED

The proposed feature selection methods are described in section III. The experimental results on vistex textures and the sidescan sonar images are reported in section IV. Finally, We draw the conclusion in section V.

II. WAVELET AND WAVELET PACKET TRANSFORM

For simplicity, a one dimensional discrete signal $f(k)$ of length $n = 2^{n_0}$ is used in this section. The wavelet transform can be thought of as a smooth partition of the signal frequency axis. First, a lowpass filter $h(m)$ and a highpass filter $g(m)$ of length M are used to filter the signal into two subbands, which are then downsampled by a factor of two. Let H and G be the convolution-downsampling operators defined as:

$$Hf(k) = \sum_{m=0}^{M-1} h(m)f(2k+m), \quad (1)$$

$$Gf(k) = \sum_{m=0}^{M-1} g(m)f(2k+m). \quad (2)$$

H and G are called perfect reconstruction quadrature mirror filters (QMFs), if they satisfy the following orthogonality conditions:

$$HG^* = GH^* = 0, \quad (3)$$

$$H^*H + G^*G = I, \quad (4)$$

where H^* and G^* are the adjoint (i.e., upsampling-anticomvolution) operators of H and G respectively, and I is the identity operator. This filtering and downsampling process is continued iteratively on the low-frequency subbands. At each level of the process, the high-frequency subband is preserved. When the process reaches the highest decomposition level, both the low- and high-frequency bands are kept. If the maximum processing level is L , the discrete wavelet coefficients of signal $f(k)$ are then $\{Gf, GHf, GH^2f, \dots, GH^L, H^{L+1}f\}$ of the same length n as the original signal. Due to the orthogonality conditions of H and G , each level of decomposition can be considered as a decomposition of the vector space into two mutually orthogonal subspaces. Let $V_{0,0}$ denote the original vector space R^n , and $V_{l,0}$ and $V_{l,1}$ be the mutually orthogonal subspaces generated by applying H and G to $V_{0,0}$. Then, the l th level of decomposition can be written as

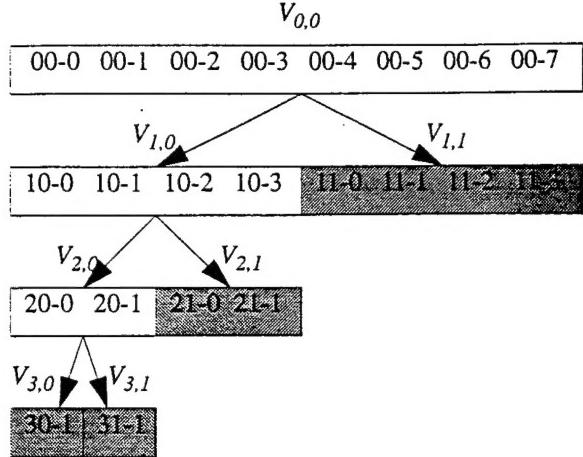


Figure 1. Standard wavelet transform binary tree.

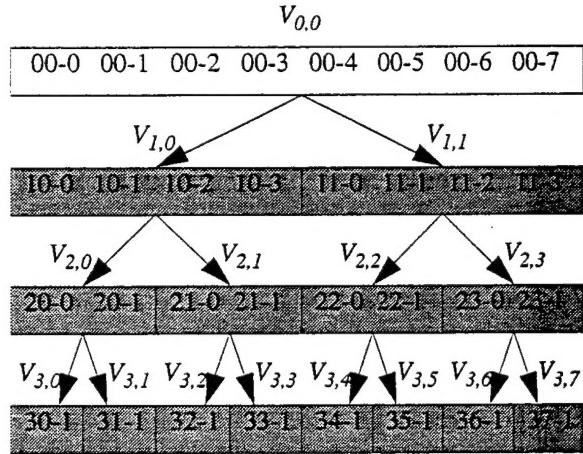


Figure 2. Wavelet packet transform binary tree.

$$V_{l,0} = V_{l+1,0} \oplus V_{l+1,1}, \quad (5)$$

for $l = 0, 1, \dots, L$. Figure 1 shows such a decomposition process. Each subspace $V_{l,b}$ with $b=0$ or 1 is spanned by 2^{n_0-l} wavelet basis vectors $\{\psi_{l,b,c}\}_{c=0}^{2^{n_0-l}-1}$, which can be derived from H , G , and their adjoint operators.

From the above iterative filtering operations, we can see that the wavelet transform partitions the

frequency axis finely toward the lower frequency region. It is suitable for a smooth signal containing primarily low frequency energy, but not necessarily appropriate for other more general types of signals, such as textures. The wavelet packet transform is a much more generalized form of the standard wavelet transform. It decomposes both the high- and low-frequency bands at each iteration. Like the wavelet transform, two subbands, Hf and Gf , are generated in the first level of decomposition. However, the second level process generates four subbands, H^2f , GHf , HGf and G^2f , instead of the two bands H^2f and GHf in the wavelet transform. If the process is repeated L times, L_n wavelet packet coefficients are obtained. In orthogonal subspace representation, the l th level of decomposition is

$$V_{l,b} = V_{l+1,2b} \oplus V_{l+1,2b+1}, \quad (6)$$

where $l = 0, 1, \dots, L$ is the level index and $b = 0, \dots, 2^l - 1$ is the channel block index in each level. Figure 2 illustrates the wavelet packet decomposition of the original vector space $V_{0,0}$. Again, each subspace $V_{l,b}$ is spanned by 2^{n0-l} basis vectors $\{W_{l,b,c}\}_{c=0}^{2^{n0-l}-1}$. For $b = 0$ and 1 , W can be identified with ψ .

As for two-dimensional images, the wavelet or wavelet packet basis function can be expressed by the tensor product of two one-dimensional basis functions in the horizontal and vertical directions. The corresponding 2-D filters are thus:

$$h_{HH}(m, n) = h(m)h(n) \quad (7)$$

$$h_{HG}(m, n) = h(m)g(n) \quad (8)$$

$$h_{GH}(m, n) = g(m)h(n) \quad (9)$$

$$h_{GG}(m, n) = g(m)g(n) \quad (10)$$

In Fig. 3, we show three sample textures and their wavelet packet coefficients for levels 1 to 4.

III. METHODOLOGY

We develop our algorithm by addressing the three main issues of multichannel texture classification: 1) feature extraction within each channel, 2) channel selection, and 3) channel relationships and feature combination among channels.

Firstly, since the wavelet coefficients are shift variant, they are not suitable for direct use as texture

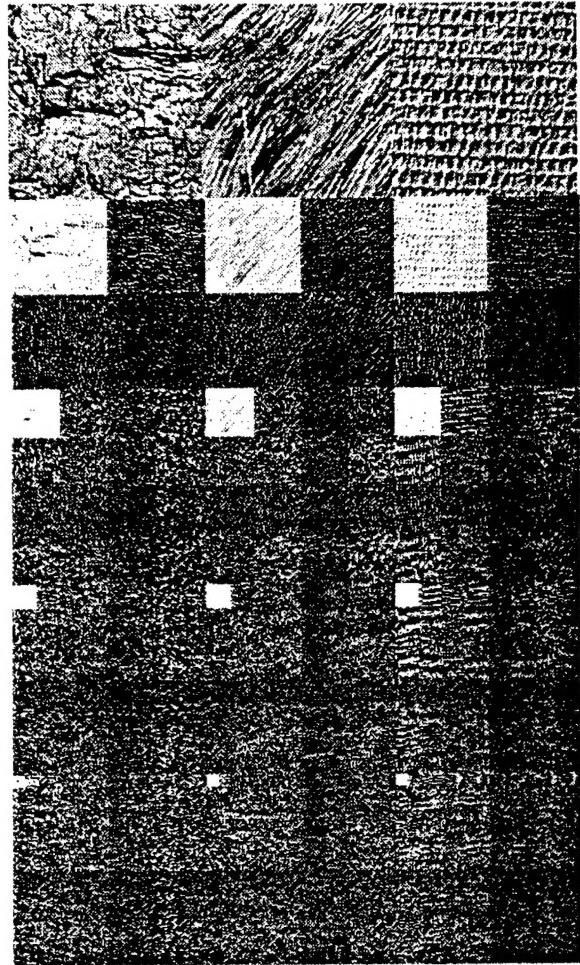


Figure 3. Three sample textures (row 1) and their wavelet packet coefficients at decomposition levels 1, 2, 3, and 4 (rows 2-5).

features. It is important to extract shift-invariant features within each channel. We choose to test the following shift invariant measurements:

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N x(i,j), \quad (11)$$

$$MNT_k = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (x(i,j) - \mu)^k, \quad (12)$$

$$ENT = - \sum_{i=1}^M \sum_{j=1}^N \frac{x(i,j)^2}{\|x\|_2^2} \log \left(\frac{x(i,j)^2}{\|x\|_2^2} \right), \quad (13)$$

where, $x(i,j)$ denotes an element of the wavelet packet coefficient matrix x in each channel. To make

our algorithm suitable for sidescan sonar images, the texture sample mean is removed before a feature vector is computed. Since the sidescan sonar image is usually cross-track range dependent even after the best effort to apply angle varying gain correction. Thus, the mean feature in equation (11) becomes zero. So the four features we use in our experiment are: 1) variance feature VAR with $k = 2$ in (12), 2) the entropy feature ENT in equation (13), 3) the third momentum MNT₃, and 4) the fourth momentum MNT₄.

Note that the orthogonality condition of the wavelet transform means that the decomposition will preserve energy. Thus for the variance feature, the following relation holds for any node and its children nodes:

$$VAR_{l,b} = \frac{1}{4} \sum_{j=0}^3 (VAR_{l+1,4b+j}). \quad (14)$$

We clearly see the effect of this linear relationship on the classification accuracy of overcomplete wavelet packet features in our experiments described in the next section.

After the features are computed within each channel, the second issue is how to select good features among channels. One possible approach is to apply a statistical distance measure to each feature and selecting the features with large distance measures. However, there are two drawbacks with this approach. The first is that neighborhood channel features tend to correlate with each other. Thus they contain similar information. If one has large distance, the other will also have a large distance measure, and both will be selected. Thus we will keep on selecting the same kind of features. The second problem is that for some very small energy channels, a small amount of unexpected noise may cause the result of a distance measure to be unrealistically large, which will be selected as a good channel. To avoid these problems, we propose to combine the channel selection step with the third step, i.e., the channel combination, into one feature selection step using the principal component analysis technique and the statistical distance measurement.

The widely used Karhunen-Loeve transform (KLT) is an ideal feature reduction and selection procedure for our algorithm. Its decorrelation ability serves to decorrelate neighborhood channel features, and its energy packing property serves to remove noisy channels and to compact useful information into a few dominant features. But for a large feature

vector, such as features in a higher level wavelet packet decomposition and the Fourier transform features, the computation of the eigenvectors of the covariance matrix could be prohibitively expensive. We use the dominant eigenvector estimation method described in [9] [11] to overcome this problem.

However, as optimal representation features, KLT selected features may not be the best for classification. Additional feature class separability measurements are used to select KLT decorrelated features. We use the Bhattacharyya distance measurement in this study.

The reason that Bhattacharyya distance is used is its direct relation to the error bound of the Bayes classifier and its simple form for features with normal distributions [11]:

$$\beta_{d(c_1, c_2)} = \frac{1}{8} (\mu_1 - \mu_2)^T \left(\frac{W_1 + W_2}{2} \right)^{-1} (\mu_1 - \mu_2) \quad (15)$$

$$+ \frac{1}{2} \ln \frac{\left| \frac{1}{2} (W_1 + W_2) \right|}{|W_1|^{1/2} |W_2|^{1/2}}$$

Because of the large number of combinations of several features and the probability of covariance matrix singularity, computing the Bhattacharyya distance for several features at once is not a practical approach. The one-at-a-time method is adopted instead. The formula is the same as Equation (15), only with covariance matrix W replaced by variance and mean vector μ replaced by class mean. As for multiclass problems, we select features with small values of

$$S_b = \sum_{i>j}^M \sum_{j=1}^M \exp [-\beta_{d(c_i, c_j)}] \quad (16)$$

In the next section we test our algorithms on the following group of features:

1. level 1: VAR, ENT, MNT₃, MNT₄, all,
2. level 2: VAR, ENT, MNT₃, MNT₄, all,
3. level 3: VAR, ENT, MNT₃, MNT₄, all,
4. level 4: VAR, all,
5. level 1&2: VAR, all,
6. level 1&2&3: VAR, all,
7. level 1&2&3&4: VAR,

8. Wavelet: VAR, all,

9. FFT: magnitude.

The goal is to test the discrimination power of each feature type in each individual level, the effect of overcomplete representation, and the classification power of the standard wavelet transform. We also test the Fourier transform features, which can be considered as an extreme case of the wavelet packet transform. i.e., the highest possible level of wavelet packet decomposition. This fourier transform feature should not be confused with the traditional power spectrum method(PSM). To compare the difference between our approach and the traditional PSM, refer to [9].

The classification algorithm used in this study is the Gaussian classifier. There are two reasons for this choice. First, it agrees with the above error bound defined by the Bhattacharyya distance. Second, with our focus on feature extraction, we choose the simplest classification algorithm available. Again, we assume the feature vector x for each class i has a Gaussian distribution with mean μ_i and covariance matrix W_i . Then, the distance measure is defined as [11]

$$D_i = (x - \mu_i)^T W_i^{-1} (x - \mu_i) + \ln|W_i| , \quad (17)$$

where the first term on the right of the equation is actually the Mahalanobis distance. The decision rule is

$$x \in C_L \quad \text{when} \quad D_L = \min \{D_i\} . \quad (18)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We test our algorithms on a set of eight types of natural optical images obtained from the MIT Media lab Vistex texture data base (see Fig. 4). The original 512x512 color images are converted to the same size gray scale images with 256 gray levels. Then, adaptive histogram equalization is applied. So all images have the same flat histogram and are indistinguishable from each other in terms first order statistics. To test the sensitivity of our algorithms to noise, we add several levels of white noise to the data. By choosing eight classes of images, doing histogram flattening, and adding noise, we try to make the classification task more difficult, so the difference in classification ability of various texture

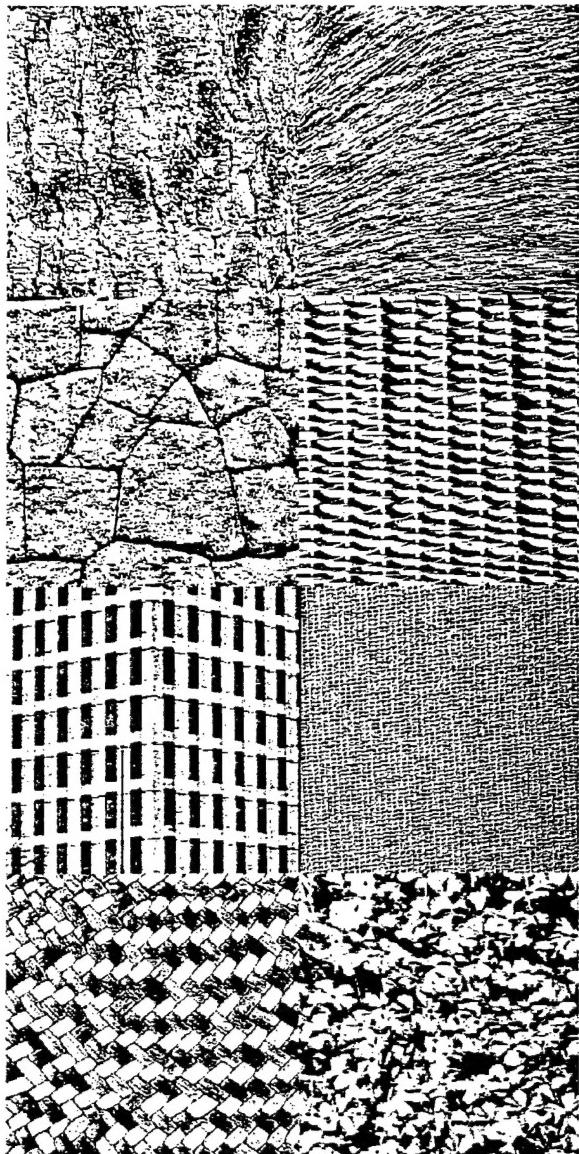


Figure 4. Vistex textures: bark.0008, brick.0004, buildings.0009, fabric.0001, fabric.0005, fabric.0013, fabric.0017, flowers.0007, from top to bottom, then from left to right.

features becomes more apparent. We then select the most successful methods to test on noisy, real-world sidescan sonar images. The three classes of sidescan sonar texture images used are shown in Fig. 5. They are first-year (young) ice, multiyear undeformed ice, and multiyear deformed ice. For the two data sets, each class of image is divided into 225 half-overlapping samples of dimension 64x64, of which 60 samples are used for training. Therefore, the total data sample number is 1800 for Vistex data and 675 for the sidescan sonar data set, with 480 and 180

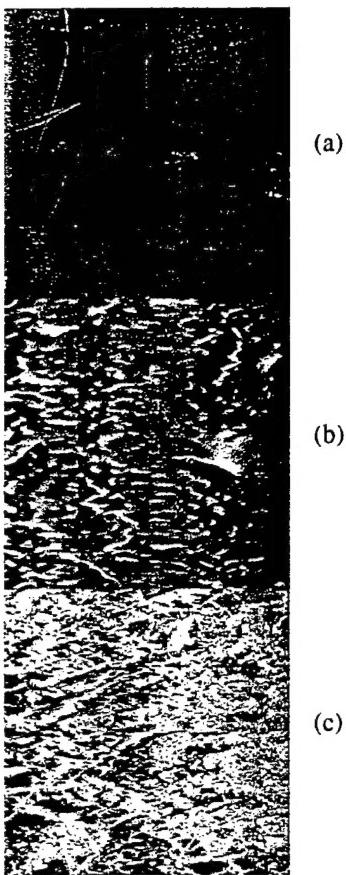


Figure 5. Sidescan sonar images of an Arctic under-ice canopy: (a) first-year (young) ice, (b) multiyear undeformed ice, and (c) multiyear deformed ice.

samples for training, respectively.

Table 1 shows the complete testing results from the Vistex data. It is somewhat overwhelming to make sense of these large amount of test results directly from this table. We point out only a few apparent features of this table, then use a few plots of the results from the table to illustrate our other findings.

First, notice that for some feature groups, the differences in classification accuracy between training and testing data are very large, more than 50% in some cases (SNR 1 data). In fact, except for the level one features, which have only four channels, almost all other training data classifications achieve more than 95% accuracy, including the SNR 1 noisy data. This is not the case for the testing data. Since only the simple Gaussian classifier is used here, we should expect these trend be even more apparent for the more sophisticated

classifier, which can learn a more precise feature structure of the training data. The significance of this result is that it shows the widely used leave-one-out testing scheme can be rather deceiving for testing new algorithms. Since leaving one sample out does not affect much of the training process. In the case of the Gaussian classifier, the effect is minimal. This means that if the data set is too small, the results will not be conclusive.

Also note in the table that the number of features used to achieve best results for each group of features is mostly about 10. The difference is not that large. A general trend is that noisier data tend to need more features to reach best classification.

To help focus on the classification accuracy of the overall data set, Fig. 6 shows the comparison of the four types of features and their combinations in the first three decomposition levels. The MNT_3 feature is the worst for all levels and all data sets. It is apparently not a useful measurement. Entropy also gives much less satisfactory results than the variance feature, and the classification accuracy drops sharply for noisy data. This contradicts the result given in [8], which shows that energy features perform only slightly better than entropy measures (within one percent). The MNT_4 feature seems to give better results than the above two features but is still less successful than the variance feature. The performance differences between the MNT_4 and the variance are consistent over all data sets and all decomposition levels, which is because they are very closely correlated features.

The observation that variance features perform better than other features is consistent with Laws' [1] experiment with features extracted from empirical frequency channels. The remaining question is whether we need other measurements to add new information. We performed a union operation on the correct classification samples of the four types of features, and the correct classification rate increased by about 5%, nearing 100% accuracy. This demonstrates that each feature has its own distinct classification power. By combining all features together, we get improved result for the lower decomposition level. Since the feature length is much smaller in these levels, an additional dimension will help more than in the higher level decomposition case. The improvement is not as impressive as the union of results. The reason is that besides the new information additional features bring in, there is also a great deal of added noise, which may overwhelm the benefit of additional features.

Table 1: Complete test results on eight Vistex texture images.

| data set | | | original images | | | | SNR 15 dB | | | | SNR 5 dB | | | | SNR 1 dB | | | |
|---------------------|------------|------------------|-----------------|---------|----------|--------------|-----------|---------|----------|--------------|----------|---------|----------|--------------|----------|---------|----------|--------------|
| feature group names | | | training | testing | all data | feature num. | training | testing | all data | feature num. | training | testing | all data | feature num. | training | testing | all data | feature num. |
| 1 | level 1 | VAR | 95.6 | 95.5 | 95.5 | 4 | 95.2 | 94.8 | 94.9 | 4 | 89.0 | 89.2 | 89.2 | 4 | 82.3 | 81.3 | 81.6 | 4 |
| 2 | | ENT | 87.7 | 86.7 | 87.0 | 4 | 86.7 | 84.3 | 84.9 | 4 | 64.4 | 58.0 | 59.7 | 4 | 49.8 | 42.7 | 44.6 | 4 |
| 3 | | MNT ₃ | 62.9 | 58.0 | 59.3 | 4 | 61.5 | 58.1 | 59.0 | 4 | 49.4 | 41.3 | 43.4 | 4 | 39.4 | 32.7 | 34.4 | 4 |
| 4 | | MNT ₄ | 89.2 | 91.1 | 90.6 | 4 | 92.3 | 91.8 | 91.9 | 4 | 89.2 | 87.5 | 87.9 | 4 | 81.3 | 80.7 | 80.8 | 4 |
| 5 | | ALL | 98.8 | 96.7 | 97.2 | 8 | 98.5 | 96.7 | 97.2 | 9 | 97.3 | 92.5 | 93.8 | 11 | 94.8 | 82.5 | 85.8 | 13 |
| 6 | level 2 | VAR | 96.5 | 96.9 | 96.8 | 6 | 96.3 | 95.8 | 95.9 | 5 | 96.7 | 93.0 | 94.0 | 9 | 97.9 | 89.8 | 92.0 | 12 |
| 7 | | ENT | 94.2 | 90.6 | 91.6 | 8 | 97.7 | 86.5 | 89.5 | 14 | 97.1 | 66.7 | 74.8 | 20 | 94.2 | 48.9 | 60.9 | 20 |
| 8 | | MNT ₃ | 80.4 | 62.3 | 67.2 | 9 | 85.2 | 66.2 | 71.3 | 11 | 96.5 | 69.8 | 76.9 | 18 | 95.8 | 56.5 | 67.0 | 20 |
| 9 | | MNT ₄ | 94.2 | 90.9 | 91.8 | 6 | 93.1 | 91.6 | 92.0 | 6 | 95.0 | 87.7 | 89.7 | 9 | 95.6 | 85.0 | 87.8 | 12 |
| 10 | | ALL | 98.8 | 97.8 | 98.1 | 12 | 99.4 | 97.7 | 98.1 | 12 | 95.8 | 90.5 | 91.9 | 9 | 97.5 | 84.1 | 87.7 | 16 |
| 11 | level 3 | VAR | 97.7 | 98.1 | 98.0 | 6 | 97.5 | 98.0 | 97.9 | 7 | 95.8 | 95.7 | 95.7 | 6 | 96.7 | 90.3 | 92.0 | 13 |
| 12 | | ENT | 97.7 | 79.2 | 84.2 | 17 | 89.8 | 75.2 | 79.1 | 11 | 96.9 | 54.8 | 66.1 | 22 | 97.5 | 40.5 | 55.7 | 26 |
| 13 | | MNT ₃ | 82.7 | 60.5 | 66.4 | 10 | 85.8 | 61.6 | 68.1 | 12 | 96.0 | 53.8 | 65.1 | 22 | 99.6 | 45.2 | 59.7 | 30 |
| 14 | | MNT ₄ | 94.8 | 92.7 | 93.2 | 7 | 94.2 | 92.0 | 92.6 | 8 | 95.0 | 91.1 | 92.2 | 9 | 97.9 | 83.9 | 87.7 | 18 |
| 15 | | ALL | 98.5 | 97.8 | 98.0 | 16 | 97.7 | 97.3 | 97.4 | 6 | 97.5 | 93.2 | 94.3 | 10 | 96.5 | 90.9 | 92.4 | 13 |
| 16 | level 4 | VAR | 97.1 | 97.5 | 97.4 | 6 | 97.5 | 96.9 | 97.1 | 6 | 96.7 | 95.9 | 96.1 | 6 | 98.5 | 94.5 | 95.6 | 12 |
| 17 | | ALL | 97.7 | 97.9 | 97.8 | 8 | 97.9 | 96.4 | 96.8 | 9 | 99.2 | 93.9 | 95.3 | 18 | 95.4 | 93.0 | 93.6 | 6 |
| 18 | level 12 | VAR | 96.5 | 96.9 | 96.8 | 6 | 96.3 | 95.8 | 95.9 | 5 | 96.7 | 93.0 | 94.0 | 9 | 97.9 | 89.8 | 92.0 | 12 |
| 19 | | ALL | 99.4 | 98.3 | 98.6 | 10 | 99.2 | 98.2 | 98.4 | 10 | 99.2 | 93.3 | 94.8 | 17 | 93.3 | 85.8 | 87.8 | 10 |
| 20 | lev123 | VAR | 98.5 | 98.6 | 98.6 | 7 | 98.3 | 98.6 | 98.6 | 7 | 96.9 | 95.3 | 95.7 | 9 | 97.9 | 90.5 | 92.4 | 13 |
| 21 | | ALL | 99.8 | 98.6 | 98.9 | 20 | 99.2 | 98.2 | 98.4 | 11 | 98.8 | 94.5 | 95.7 | 13 | 97.7 | 91.7 | 93.3 | 14 |
| 22 | lev1234var | VAR | 99.0 | 97.2 | 97.7 | 14 | 98.8 | 96.7 | 97.3 | 12 | 97.9 | 96.0 | 96.5 | 12 | 98.1 | 93.7 | 94.9 | 11 |
| 23 | wavelet | VAR | 98.1 | 97.0 | 97.3 | 10 | 97.9 | 96.7 | 97.1 | 10 | 97.7 | 95.0 | 95.7 | 11 | 97.9 | 94.2 | 95.2 | 12 |
| 24 | | ALL | 100 | 96.6 | 97.5 | 18 | 100 | 96.4 | 97.4 | 18 | 97.9 | 93.8 | 94.9 | 14 | 98.1 | 90.2 | 92.3 | 19 |
| 25 | FFT Mag. | VAR | 99.8 | 98.4 | 98.8 | 26 | 99.4 | 97.9 | 98.3 | 15 | 99.0 | 96.4 | 97.1 | 10 | 99.8 | 95.8 | 96.8 | 17 |

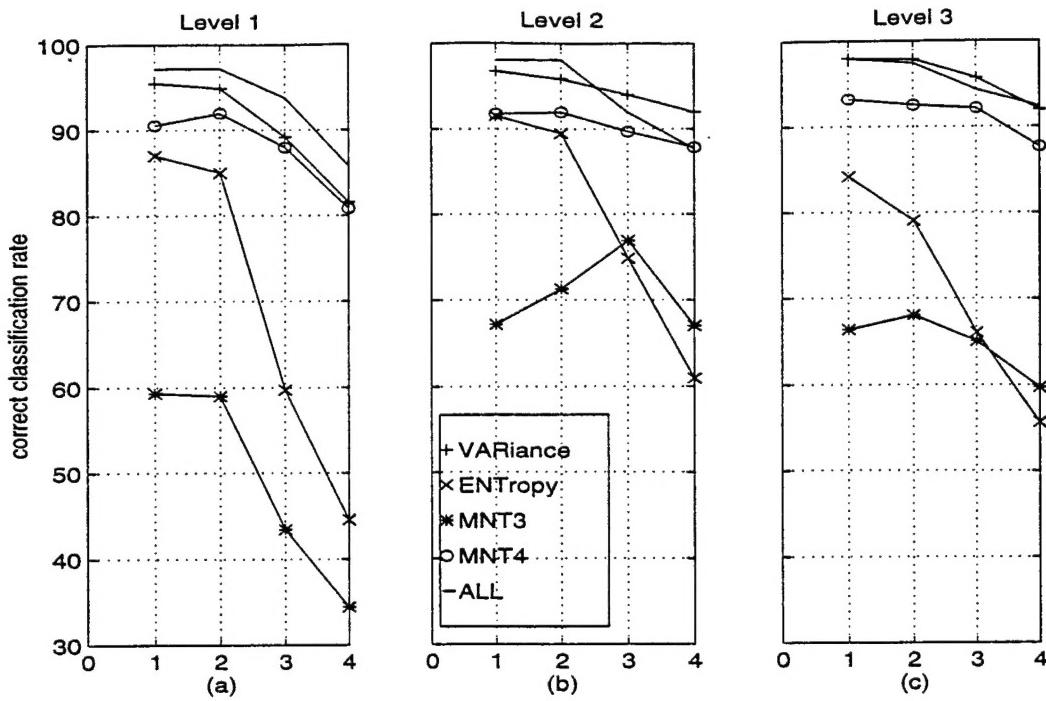


Figure 6. Comparison of the four types of features in the first three individual decomposition levels. The index of the horizontal axis represent signal-to-noise ratio (SNR) level: 1. original images, 2. SNR 15dB, 3. SNR 5dB, 4. SNR 1dB.

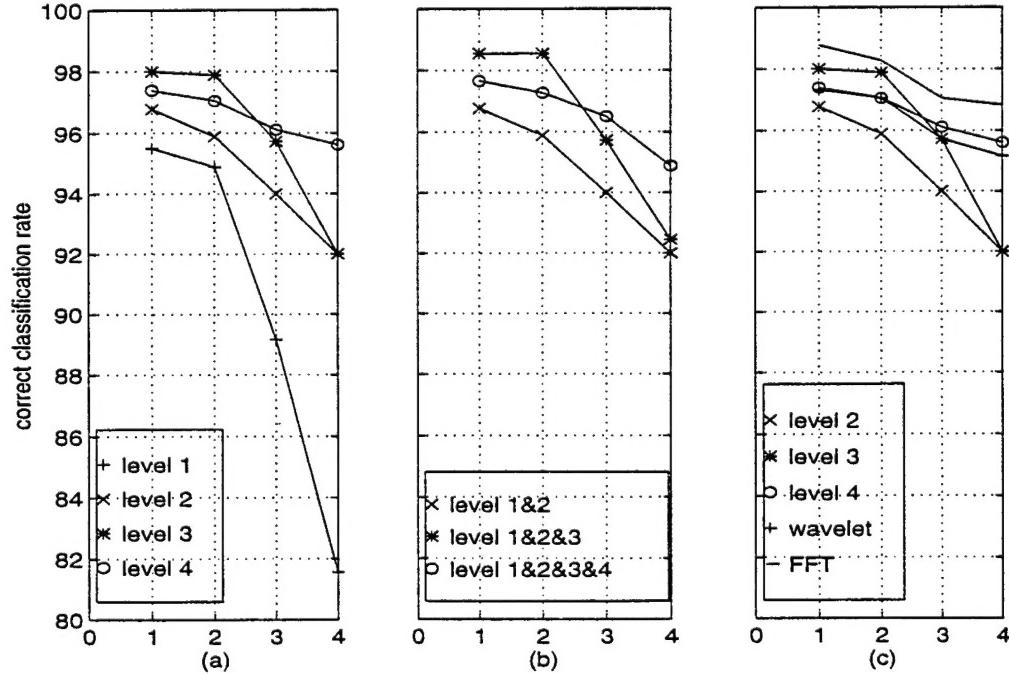


Figure 7. Comparison of variance features for individual decomposition levels, overcomplete levels, standard wavelet, and Fourier transform. The index of the horizontal axis represents the same SNR as in Fig. 6.

We now look in detail at the variance measurement results shown in Fig. 7. From Fig. 7(a), for individual levels, the general trend is the higher the decomposition level the better the result. This is predictable from Equation (14), which shows that the lower level variance features are simply the average of their higher level children nodes. A KLT transform will do better than such a simple average operation in terms of extracting maximum information. To confirm this point, compare Fig. 7 (a) and (b). It is easily seen that the following pairs of results are almost identical: level 1&2 vs. level 2, level 1&2&3 vs. level 3, level 1&2&3&4 vs. level 4. This means that lower lever features are only a subset of higher level decomposition features. This is contrary to what Laine and Fan suggested in [8], that redundancy may provide additional discrimination power. Our experiments show that better discrimination ability is not added by overcompletion. Instead, it is extracted by applying KLT to higher levels of finer channel decomposition, so the channel nodes are combined in an optimal way, instead of by simple averaging.

Continuing further along this path, we should expect that the Fourier transform may provide even more information. Because the Fourier transform is really the extreme case of a wavelet packet transform, i.e., the wavelet packet transform at its highest possible level. Figure 7(c) compares the performances of three levels of wavelet packet decomposition, the standard wavelet transform, and the Fourier transform. The Fourier transform indeed gives a consistently better performance over all other feature groups on all levels of noisy data sets. This result should not come as a surprise, since the wavelet transform is optimal for nonstationary signal analysis, whereas the Fourier transform is optimal for stationary signal analysis. Most texture images are stationary periodical signals.

Next, notice in Fig. 7(c) that the Fourier transform and other higher levels of wavelet packet decompositions are very insensitive to noise. It is surprising to see that more than 95% accuracy is achieved at a noise level of 1dB, compared with the results in [7], where the tree-structured wavelet algorithm collapses to 70% accuracy at a 5dB noise level. Noise insensitivity is really the strength of subband image processing. Noise usually has a flat spectrum, so by dividing into more subbands, the noise energy usually decreases. Yet the energy of signals tend to concentrate in a small number of channels. Therefore, even when the total energy of the signal and noise are almost the same, as in the case of our testing data of SNR 1dB, the signal-to-

noise ratio will be much higher in channels containing most of the signal energy. Our feature selection algorithms are designed in such a way that they pick up and condense the signal channel with high SNR into a compact representation of the data, with the incoherent noisy channel neglected.

Finally, we test our algorithms on the classification of sidescan sonar images. Only the feature groups that performed best in the above experiment are used on the sonar images. Table 2 shows the results, which are consistent with the Vistex data results. Although the image class number is smaller, each class of image is noisy and nonuniform. This added difficulty increases the accuracy difference between the wavelet packet and Fourier transform methods. It again shows the superiority of the latter approach.

Table 2: Classification results on sidescan sonar images.

| feature group names | | training | testing | all data | feature num. |
|---------------------|----------------|----------|---------|----------|--------------|
| 1 | level 3 VAR | 93.3 | 83.2 | 85.9 | 12 |
| 2 | level 4 VAR | 96.1 | 81.4 | 85.3 | 20 |
| 3 | all levels VAR | 98.3 | 85.3 | 88.7 | 23 |
| 4 | stand. wavelet | 88.3 | 88.1 | 88.1 | 7 |
| 5 | FFT Mag. | 98.9 | 92.5 | 94.2 | 9 |

V. CONCLUSIONS

Based on the above experiments, the following conclusions are drawn:

- 1). Variance (energy) measurement is much better than entropy and higher order momentum. But there does exist additional information in the latter three features that is distinct from the energy information. Better feature selection algorithms have yet to be developed to fully employ this information.
- 2). Higher levels of decomposition perform better than lower levels. This leads to the conclusion that Fourier transform features are better than wavelet packet features.
- 3). For variance features, overcomplete

representation does not give better results than individual level features.

4). Wavelet packet features are very insensitive to noise. Features from higher levels are less sensitive than the lower level wavelet features. This noise insensitivity property makes wavelet packet features suitable for sidescan sonar images, which are usually noisy.

5). The KLT and Bhattacharyya distance measurement methods are good feature selection methods to use on wavelet packet features. KLT is necessary, especially for higher level features.

6). There are great differences between the training data and testing data classification accuracy. It casts doubt on results of the leave-one-out testing strategy used by many texture classification works.

ACKNOWLEDGEMENTS

The sidescan sonar images used in this study are produced at the Deep Submergence Lab of the Woods Hole Oceanographic Institution, and many individuals have participated in the development and application of the sonar-processing techniques. In particular, we wish to thank M. Marra, H. Singh, S. Lerner, and D. Chu for their efforts. We would also like to thank the MIT Media Lab for providing the Vistex texture images. This work is supported by the Office of Naval Research through grant N00014-93-1-0181. This is contribution number 9031 of the Woods Hole Oceanographic Institution.

REFERENCES

- [1] K. I. Laws, "Textured Image Segmentation," Ph.D. dissertation, Image Processing Inst., Univ. of Southern California, 1980.
- [2] M. Unser and M. Eden, "Multiresolution feature extraction and selection for texture segmentation," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 11, no. 7, pp. 717-728, July 1989.
- [3] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," in *Proc. of the IEEE Inter. Conf. on Sys. Man and Cyber.*, pp. 14-19, Los Angeles, CA, 1990.
- [4] S.G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 11, pp. 674-693, July 1989.
- [5] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713-719, Mar. 1992.
- [6] M. B. Henke-Reed and S. N. C. Cheng, "Cloth texture classification using the wavelet transform," *Journal of Imaging Science and Technology*, vol. 37, pp. 610, Dec. 1993.
- [7] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Processing*, vol. 2, pp. 429-441, Oct. 1993.
- [8] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 15, pp. 1186-1191, Nov. 1993.
- [9] X. Tang and W. K. Stewart, "Texture classification using principle-component analysis techniques," *Proc. Image and Signal Processing for Remote Sensing*, Rome, Italy, 1994.
- [10] A. Goff, W. K. Stewart, H. Singh, and X. Tang, "Quantitative analysis of ice draft II: application of stochastic modeling to intersecting topographic profiles," *J. Geophys. Res.*, vol. 100, No. C4, pp. 7005-7017, April, 1995.
- [11] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, New York, 1972.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

| | | | | | | | |
|---|--|--|---|---|---|-----------------------------------|--|
| 1. AGENCY USE ONLY (Leave blank) | | | 2. REPORT DATE 1995 | | 3. REPORT TYPE AND DATES COVERED TECHNICAL | | |
| 4. TITLE AND SUBTITLE Texture classification using wavelet packet and Fourier Transforms | | | 5. FUNDING NUMBERS N00014-93-I-0181 | | | | |
| 6. AUTHOR(S) Xiaou Tang and W. Kenneth Stewart | | | 8. PERFORMING ORGANIZATION REPORT NUMBER WHOI CONTR. 9031 | | | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) WOODS HOLE OCEANOGRAPHIC INSTITUTION WOODS HOLE, MA 02543 | | | 10. SPONSORING/MONITORING AGENCY REPORT NUMBER | | | | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) OFFICE OF NAVAL RESEARCH ENVIRONMENTAL SCIENCES DIRECTORATE ARLINGTON, VA 22217-5660 | | | 12a. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED | | | | |
| 11. SUPPLEMENTARY NOTES In citing this report in a bibliography, the reference given should be: Oceans 95, I:387-396, 1995. | | | 12b. DISTRIBUTION CODE | | | | |
| 13. ABSTRACT (Maximum 200 words) | | | | | <i>Abstract —A new texture classification algorithm using wavelet packet transform is proposed. It uses principal component analysis technique and statistical distance measurement to combine and select frequency channel features to give improved classification performance. Comparison is also made between wavelet packet transform features and Fourier transform features on a set of eight optical texture images with several level of white noise added. Both algorithms are successfully applied to the classification of under-ice sidescan sonar images.</i> | | |
| 14. SUBJECT TERMS | | | 1) texture classification 2) wavelet packet transform 3) Karhunen-Loeve transform | | 15. NUMBER OF PAGES 10 16. PRICE CODE | | |
| 17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED | | 18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED | | 19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED | | 20. LIMITATION OF ABSTRACT | |